Temporal evolution of communities based on scientometrics data

Matteo Morini, Pablo Jensen, Patrick Flandrin

Temporal Evolution of Communities Based on Scientometrics Data

Matteo Morini, Pablo Jensen & Patrick Flandrin
LIP, LP, IXXI, ENS de Lyon

[PRELIMINARY VERSION]

1. Motivation and methods summary

History of science is made of both evolutions within (or at the boundary of) known disciplinary territories and, from time to time, of abrupt changes corresponding to the emergence of new concepts or tools, that may eventually reshape the landscape. With the development of comprehensive bibliographic databases, scientific developments can nowadays be explored in detail in the publication records, thus providing new insights that complement more traditional historical or sociological approaches.
Wavelets offer a paradigmatic example of the emergence of a new concept/tool/approach/idea, that significantly impacted many areas, from theory to applications. Studying the history of its emergence is interesting for several reasons. First, wavelets are of particular relevance for signal/image processing, which benefited from new bridges between mathematics, physics and electrical engineering. On a more methodological ground, wavelets history is interesting because it is a recently born subfield (seminal paper in 1984), for which most actors are still around for feedback. Moreover, there exists a large corpus of publications which allow for a meaningful use of our quantitative tools.

To retrieve a relevant set of publications, we identified 83 key actors of early developments of wavelets[1]. We then retrieved *all* their publications (from 1970 to 2012), obtaining 6,500 records from Web of Science [1].

2. Key results

a. Summary

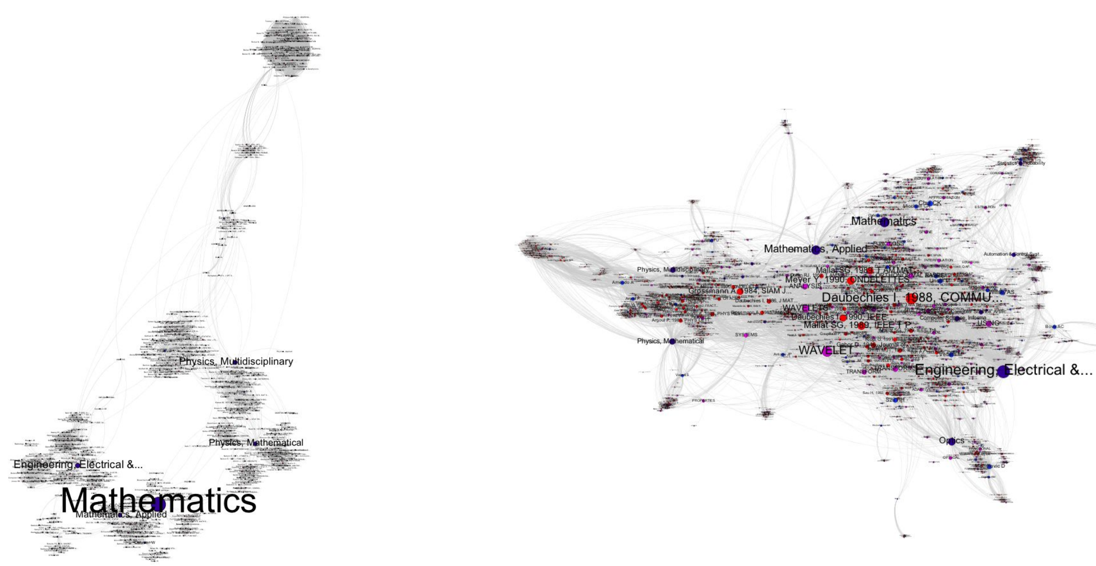
Our analysis shows three main stages. In an initial phase (before ~ 1985), researchers work in different, relatively unrelated fields. Then, wavelets appear as a common ground/identifier whose use gains momentum, defining a new, specific field that interlinks scholars. Finally (after ~ 1993), wavelets become a mature tool, that are less an object of interest *per se*, serving

[1] The list was established using expert advice (one of the authors (PF) and Alain Arneodo) and bibliographic searches. An important ingredient was the comprehensive *Wavelet Literature Survey* [2] published in 1993, a year that can be considered as a turning point in the development of the field, with, e.g., the appearance in the January issue of the *IEEE Transactions on Signal Processing* of “Wavelets and Filter Banks” as entry SP 2.4.4 of its EDICS and, in December, the publication of a Special Issue on “Wavelets and Signal Processing” in the same journal.

instead a more ancillary role within specialized communities and paving the way for new avenues of research. The rest of this article details these findings.

b. Two snapshots, two key moments

The two first stages can be clearly visualized using heterogeneous graphs. In these graphs, the authors, references, keywords and disciplines of the articles are represented as nodes, which are linked whenever they appear together in the same article. Spatializing these networks with force-based algorithms provides a relevant visualization of the structure of the scientific fields [3]. Indeed, Figure 1a shows several weakly connected components, corresponding to different authors (and their associated subfields) which share only few elements. As expected for a collection of scientists working on different topics, the graph shows a set of almost disconnected “islands”. Ten years later (Figure 1b), the landscape has changed dramatically: The articles of the *same* authors are organized in a strongly connected community, with very important networking references such as (Daubechies, 1988) or (Grossmann & Morlet, 1984).
In the remaining section, we investigate how one goes from Figure 1a to Figure 1b, and what has happened since.



a (1980-83) *b (1991-1994)*

Figure 1: Heterogeneous network, mixing authors, keywords, references cited and disciplines. The size of the labels is proportional to the number of articles of our database in which an item appear. The width of the links indicates the co-occurrence weight between two items (we keep only links corresponding to two or more co-occurrences). Colors correspond to the type of the items (authors in blue, keywords in pink,

disciplines in purple and references in red). Zoomable versions of these Figures are available at http://perso.ens-lyon.fr/matteo.morini/wavelets/mapszoom/map1980_1983.pdf (a) and http://perso.ens-lyon.fr/matteo.morini/wavelets/mapszoom/map1991_1994.pdf (b)

c. *The history*

To study in detail the transition sketched in Figure 1, we have grouped the articles in 4-years wide time slices, separated by one year (therefore, there is a 3-year overlap between subsequent time slices). For each slice, we find the relevant scientific structure (topics) by standard scientometrics methods. We first define a network of the articles of the time slice by linking the articles that share references. We then compute the relevant communities by a recursive modularity optimization [5, 6]. Figure 2 displays the time evolution of these communities. For this, we link subsequent communities according to their overlap using a Jaccard index [7]. We then obtain the genealogy of the communities structuring the wavelet history from its birth to its more recent avatars. Figures 3 and 4 zoom in on important moments: The first convergence related to Daubechies' work (electrical engineering meets mathematics, Figure 3) and the renewal of wavelet theory in relation with a growing interest for sparsity and the emergence of "compressed sensing" as a field of its own (Figure 4).

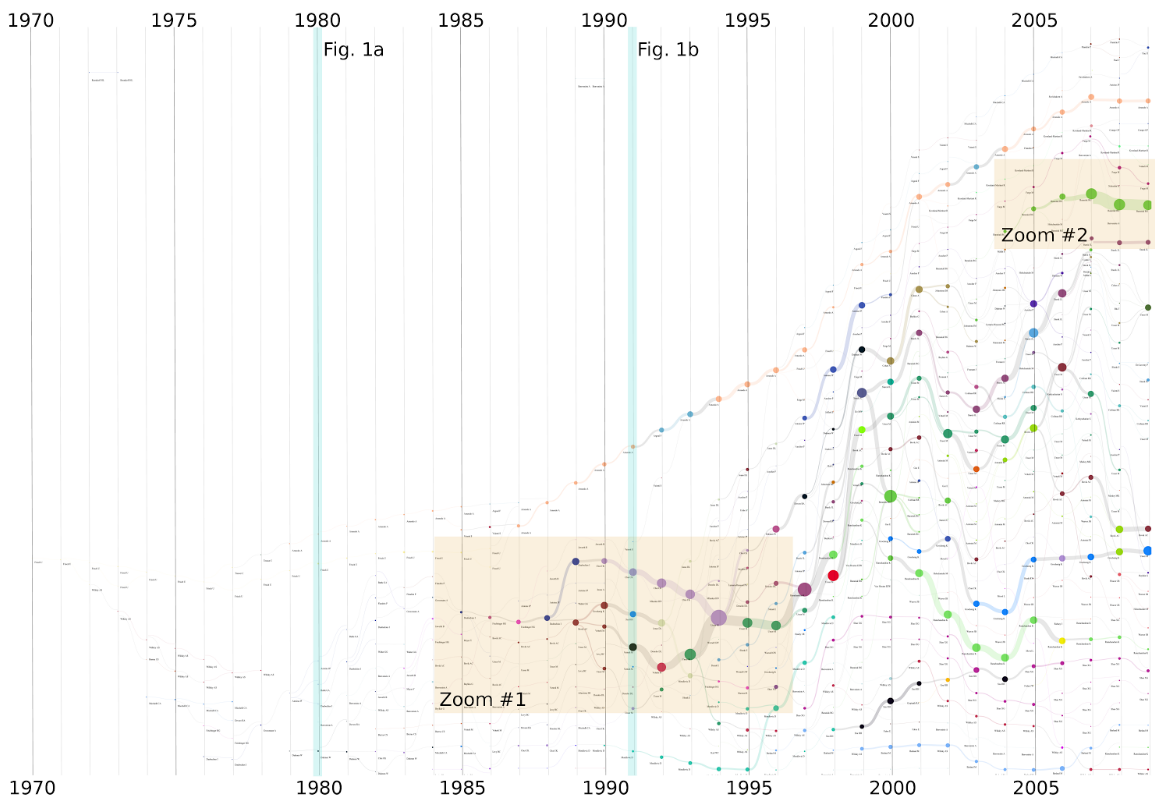


Figure 2: Overview of the alluvial diagram. Nodes are labelled according to the most frequent (top) author in each cluster. Zoomable version available at http://perso.ens-lyon.fr/matteo.morini/wavelets/sankeyzoom/sankey_s.pdf

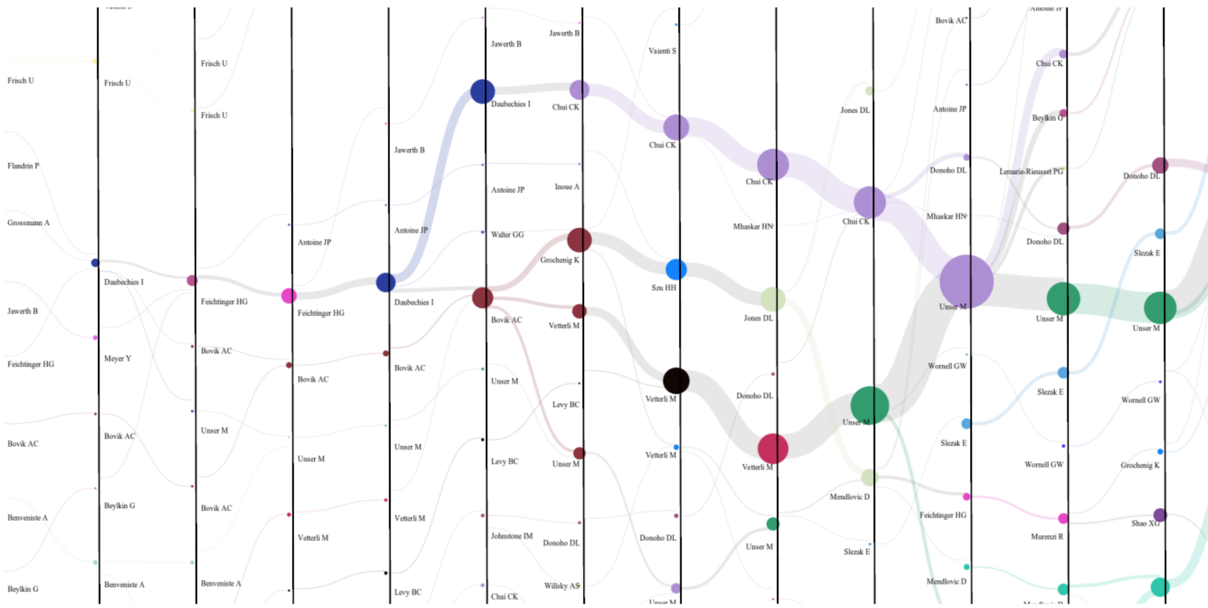


Figure 4: Clusters condensing around the inception of “wavelets” (photo1/zoom #1)

Figure 5: Cluster size dynamics, compressed sensing (top author: RG Baraniuk) gaining momentum (photo2/zoom #2)

The convergences/divergences observed in Figure 2 can be quantified by computing the *modularity* [5] of the network obtained at each time slice. Roughly speaking, a high modularity value corresponds to isolated clusters, while low values point to highly interconnected networks. Figure 6 then confirms the progressive interlinkage of the isolated communities starting in 1984 and ending in ~1994, where modularity reaches its minimum (0.45). After this, modularity increases again, pointing to a new, softer divergence, as the initial levels are not reached. This can be interpreted as a divergence of the different authors, which now are no longer so strongly related by a common interest in developing wavelets as a tool, but rather by developing new tools (as exemplified by the emergence of “compressed sensing”) or applying wavelets to specific, relatively unrelated, domains.

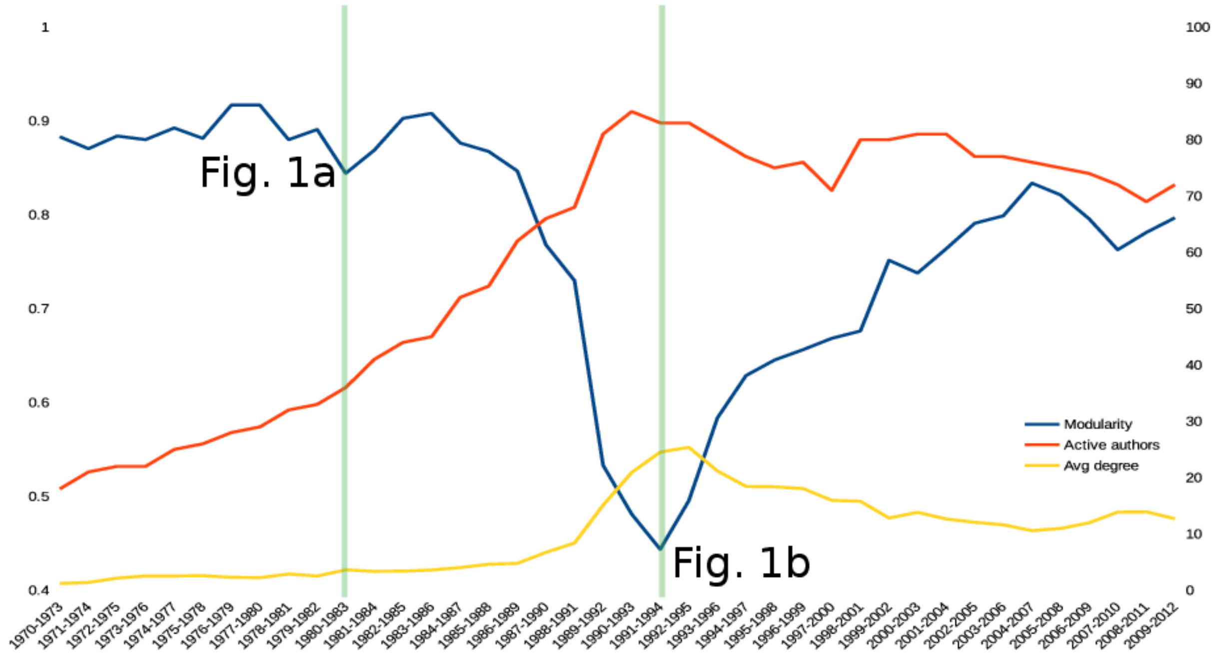


Figure 6: evolution of the modularity of the wavelets communities. A high value corresponds to isolated communities, while low values point to highly interconnected networks. The green lines point to the time slices in Figs 1a/b. In red and yellow, respectively, plotted against the right hand y axis, the number of active authors, and the average degree of each time slice.

3. Concluding remarks

We have only briefly sketched our story on the birth and evolution of the “wavelets paradigm” from bibliometric data. The purpose of the proposed article is to show in more detail the history, elaborate on the method and propose some interpretations about our findings. Our scope is to present a clear and objective story that can be read by all scientists working with wavelets. We also seek feedback from the scientists that participated in this adventure, to favor further investigations and improve our tool for semi-blind history.

References :

[1] http://apps.webofknowledge.com/ (restricted access)

[2] S. Pittner, J. Schneid, and C.W. Überhuber, *eds*. (1993). *Wavelet Literature Survey*, Technical University Vienna, ISBN 3-901363-00-9.

[3] M. Bastian, S. Heymann, and M. Jacomy (2009). *“*Gephi: an open source software for exploring and manipulating networks*”*, International AAAI Conference on Weblogs and Social Media.

[4] S. Grauwin, G. Beslon, É. Fleury, S. Franceschelli, C. Robardet, J.-B. Rouquier, and P. Jensen (2012). "Complex systems science: Dreams of universality, interdisciplinarity reality", *Journal of the American Society for Information Science and Technology*, Vol. 63, No. 7, pp. 1327-1338

[5] M.E.J. Newman (2005). "A measure of betweenness centrality based on random walks", *Social Networks*, Vol. 27, No.1, pp. 39–54.

[6] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 10, P10008-19.

[7] P. Jaccard (1901). "Etude comparative de la distribution florale dans une portion des Alpes et du Jura", *Bulletin de la Société vaudoise des Sciences Naturelles*, Vol. 37, pp. 547–579

Addendum: List of WLTC selected authors

Aldroubi, A; Antoine, JP; Antonini, M; Argoul, F; Arneodo, A; Auscher, P; Bacry, E; Baraniuk, RG; Barlaud, M; Battle, GA; Benveniste, A; Beylkin, G; Bijaoui, A; Boudreaux-Bartels, GF; Bovik, AC; Burrus, CS; Chui, CK; Cohen, A; Coifman, RR; Compo, GP; Dahmen, W; Daubechies, I; Devore, RA; Donoho, DL; Elezgaray, J; Escudie, B; Farge, M; Feauveau, JC; Feichtinger, HG; Flandrin, P; Frisch, U; Froment, J; Gopinath, RA; Grochenig, K; Grossmann, A; Guillemain, P; Haar, A; Healy, DM; Heil, C; Herley, C; Holschneider, M; Jaffard, S; Jawerth, B; Johnstone, IM; Jones, DL; Kadambe, S; Kaiser, JF; Kerkyacharian, G; Kronland-Martinet, R; Lawton, W; Lemarie-Rieusset, PG; Lu, J; Lucier, BJ; Mallat, S; Mendlovic, D; Meyer, Y; Micchelli, CA; Morlet, J; Murenzi, R; Muschietti, MA; Muzy, JF; Paul, T; Perrier, V; Picard, D; Ramchandran, K; Resnikoff, HL; Rioul, O; Saracco, G; Shao, XG; Shapiro, JM; Slezak, E; Starck, JL; Szu, HH; Tchamitchian, P; Torresani, B; Unser, M; Vaienti, S; Vetterli, M; Walter, GG; Weaver, JB; Wickerhauser, MV; Willsky, AS; Wornell, GW.